The Ruth H. Hooker Research Library

and Technical Information Center

OPTICAL IMAGE TECHNOLOGY USED FOR INFORMATION MANAGEMENT BY THE LIBRARY OF THE NAVAL RESEARCH LABORATORY

Laurie E. Stackpole, Chief Librarian,
Naval Research Laboratory, Washington, D.C. 20375-5334
Doris R. Folen
Kestrel Associates, Arlington, VA 22206

- **Abstract:**

The Ruth H. Hooker Research Library and Technical Information Center of the Naval Research Laboratory (NRL) in Washington, D.C. has installed an optical disk system for the storage and retrieval of its very large collection of technical report literature. The reports are stored as optical images on 12-inch 6.55 gigabyte disks housed in a 50-platter Sony jukebox. The reports are fully indexed in a Cuadra Star database resident on an AlphaMicro 3000 minicomputer and the optical images are linked to this system. The images are retrieved by using the ImageExtender capability of Cambridge imaging software. 'Me library patron uses a 486 PC viewing station with full-page monitor to search the database, identify, retrieve and view needed reports. All or part or the retrieved report can be printed out by the user on a high speed printer. Plans are underway to scan in other types of library material including journal articles, and to expand this system to allow retrieval via the campus network giving the NRL scientists the option of convenient remote access to library material from their laboratories or offices.

- **Introduction**

Historically, libraries were among the first of the information managers. It is known that the great library of Alexandria which flourished over twenty two hundred years ago successfully managed the largest collection of information then in existence: collecting, storing, organizing and retrieving thousands of written records.

Collecting, storing, organizing, retrieving and delivering are still among the major components of successful information management. That much has not changed over the millennia. What has changed are the tools to implement these tasks and the methods of getting the information to those who need it.

Optical image technology has the potential to revolutionize the traditional operation of a library and to change the role of libraries in the storage and dissemination of information.

This paper addresses the optical disk storage capability at the Ruth H. Hooker Research Library and Technical Information Center at the Naval Research Laboratory (NRL), Washington, D.C. This technology brings to reality a scenario in which a librarian, an information specialist or a scientist can sit in an office environment, identify a paper, document or picture and retrieve the item in minutes without leaving the area.

- **The NRL Environment**

NRL was established in 1923 by the Government at the urging of Thomas Edison for a National laboratory to serve the country's defense needs. Today NRL serves as the Navy's corporate laboratory, conducting research in such areas as acoustics, biochemistry, chemistry, electronics, information technology, materials science, particle and plasma physics, optics and space science. The Library meets the information needs of NRL's research community located at the Washington, D.C. campus in Southwest Washington, D.C. This user group consists of about 3,500 Federal staff and about 1500 contractors.

- **The Research Reports Collection**

  The optical disk system is located in the Research Reports Section of the Library. The Research Reports Section was created in 1945 when it was thought that the end of World War 11 meant the end of the production of reports by government agencies and by government contractors. The existing reports at NRL were gathered together to form an archive collection that, it was believed, once organized would not grow very much. However, the opposite proved to be the case. 'Me production of technical reports continued to accelerate during the following decades. At this date over 600,000 titles have been accessioned and cataloged by the Section. Over the years the individual collections have been reduced so that 300,000 hard copies remained on the shelves.

  These reports are considered invaluable and irreplaceable, dating back to the early 40's with some reports even older. They cover research in the areas of physics, chemistry, oceanography, electronics, metallurgy, optics, engineering, intelligence, artificial intelligence, etc. They originate from research facilities located at government installations, universities and corporate laboratories not only in this country but also abroad.

  The first segment of this collection to be put on optical disk is a group of 140,000 reports averaging 55 pages each. These reports have been issued by the Naval Research Laboratory, by other government agencies, by government contractors and by universities. Over one-third of the collection, about 60,000 reports or 3.3 million pages, have already been scanned.

- **The Optical Disk System configuration**

  Fig. 1 shows the basic hardware that constitutes the system. It was designed to house the entire reports collection and will be able to do so with the addition of a second autochanger.

  The reports are scanned and their images compressed on the TDC 4530 scanner. The images are then placed on the optical disks via the "cruncher," verifier and optical server, which are all 486 PCs.

  Images are retrieved at the viewing stations by first searching the Cuadra Star database resident on the AlphaMicro minicomputer. Once a desired report is identified, its image is displayed by placing the cursor over the report's unique identifier and entering a single keystroke.

  The Sony Writable Optical Disk Autochanger Model WDA-610 (Fig. 2) is the nucleus of the system. It is designed for use with Sony's 12-inch optical disks. The autochanger has a footprint of eight square feet and looks from the front remarkably like a two door refrigerator. To be consistent with the market, Sony now calls its autochanger a "jukebox," which it resembles in operation.

  One 12-inch "write once read many" (WORM) optical disk can hold 6.55 gigabytes (GB) of digital data, enough to store the contents of 130,000 typewritten pages. Housing 50 disks, one autochanger provides the equivalent storage space of 500 file cabinets. Four autochangers can be daisy chained to expand storage capacity to 1.310 terabytes of on-line data on a single SCSI interface.

  A TDC DocuScan DS-4530 scanner has been incorporated into the system as the primary scanner. This scanner is equipped with a sheet feeder and a monitor. It is capable of sustained scanning at a rate of 40 sheets a minute and can scan both sides of a sheet at the same time. Scanning is being done at 300 dpi. After enhancing the image electronically, the scanner compresses the image into the industry standard CCITT Group IV format.

  When a stored image is to be viewed, it is retrieved via the file server, decompressed, and either read at a work station or printed out on an HP Laser-Jet printer capable of printing 17 pages/min. The equipment is networked internally on a Novell LAN and can be connected to the campus wide network, which is part of the Internet.

  The software used for storage and retrieval of the images is the Cambridge program with its ImageExtender capability, which links the image with the existing Star database catalog.

All bindings, staples, etc. that hold a report together must be removed before scanning. Reports can be scanned more readily if the edges are fairly even, and use of the automatic document feeder requires that edges be even. For "exploding" reports and cutting even edges, a heavy duty paper cutter (Fig. 3) is used. It has a height of 54 inches, a footprint of 3 X 4 feet, and weighs 530 pounds. It requires two hands on the control panel for operation to keep the operator out of harm's way.

A 18KVA Uninterruptible Power Supply was installed to provide power in the event of a general blackout. This system provides 14 minutes of full load power or 39 minutes half load which gives the entire computer operation time to take itself down in an emergency without any loss of data.

A comprehensive back-up routine has been devised consisting of daily tape backup-ups of everything that was scanned that day, disk-to-disk back-up when a optical platter has been filled, and a back-up of all software on Exabyte tape.

- **The Conversion Process**

The reports that are scanned take a one way trip to the scanner. The quality of the images is consistently good enough to allow the destruction of the reports after they have been scanned. There are, of course, some exceptions. Occasionally there is a report that is of intrinsic historical value so the original is preserving. These few reports are scanned and then the originals are saved in an historical file.

Because this entire collection is to be scanned, no particular care has been taken with scanning order. Reports that have come back from circulation are put aside to scan because there is no sense filing them and then taking them off the shelf later to scan. New NRL originated reports are scanned as soon as the Library receives a copy. Reports are taken from the shelves in sequence but are not checked for misfiling as they are prepared for scanning because, unlike paper reports that are filed on a shelf, images do not have to be placed on the disk in any particular order.

Time spent in preparation is well invested. Duplicate reports are removed. Reports that are oversized or of really poor quality are put aside for later consideration.

A report is prepared for the scanner by:

1. Removing all blank and extraneous pages especially covers when possible. Usually all of the information on the cover is repeated on the title page. The back covers seldom have anything on them at all;
2. For in-house scanning the reports are then "exploded" by a quick cut by the paper cutter. This removes all binding, staples, etc. If necessary, additional trimming of the sides is done also;
3. Oversize pages and fold-outs are cut to 8 1/2" x 11 " or less;
4. Loose color pictures are removed and filed.

It is important to establish a routine for the work flow if there are a lot of reports to be scanned (Fig. 4). An occasional report to be scanned is one thing. When there are thousands of them to be reduced to images, efficiency becomes the watchword.

Use of the TDC DS-4530 scanner is almost fully automatic. Sheets are sent by a document feeder through the scanner at a rate of 40 sheets per minute. Bar coded accession numbers are read by the scanner and entered. The bar code also signals the first page of the next document. Quality of the scanned image is checked page by page on a preview monitor by the operator as the pages are scanned. Also various internal electronic checks are made to insure quality, i.e. checks for skewing, compression errors, etc. A 486 PC displays the progress of the document as it is scanned and the images compressed. It also records scanning statistics.

- **Retrieval**

Reports are retrieved by the user at a viewing station. 'Me user searches the Star data base (Fig. 5). When the search is complete the user views the search results (Fig. 6). Once a report is identified it can be recalled by

placing the cursor in the "Document Number" field and pressing a function key. The report originally appears on the screen with the search results (Fig. 7). The user can view the report this way or the screen can be changed to display only the report.

The report can then be viewed page by page on the screen or the user can skim through the report or page forward or back at any desired pace. A page can be enlarged for detail or reduced in size to allow other pages to be displayed on the screen. All or part of the report can be printed out by a simple command. More than one report can be viewed on a screen and the same report can be viewed simultaneously at more than one work station at the same time.

The ability to identify the specific document to be retrieved has become a stumbling block for many system designers. Hundreds of thousands of pages of records placed on optical disk can rapidly submerge individual items in a ocean of data. Some designers solve this problem by superimposing an indexing system directly on the optical disk, but this takes both time (for indexing) and disk space (for storing index terms). Some designers superimpose ASCII text that can be searched word by word directly on the disk, but this again takes time (for OCRing) and disk space (for storing the ASCII text). Selecting an indexing system can take as much or more time as selecting the hardware for the Optical Disk System. An indexing program can be expensive and if the wrong one is chosen very costly indeed; if the cataloging information is superimposed on the optical disk, it will be very difficult to change.

The NRL Library finessed this retrieval dilemma by opting to keep the retrieval system entirely separate from the Optical Disk System. The reasoning was:

1. An automatic retrieval system was already in place;
2. There was a need for the quickest and the most expedient means of getting the reports to optical disk and a separate retrieval system supported that goal;
3. A separate system is less expensive and easier to manage, input can be done at its own pace, and anything on the system can be changed at any time with little fuss;
4. At the beginning of the planning stage the optical disk system was thought of as simply an alternative way of storing reports;
5. With the right imaging software, the fact that the cataloging database is separate from the image file can be made transparent to the user.

Since August of 1987, the Library has been using the Cuadra Star retrieval system as its reports catalog. This system provides a very large number of fields so that a report can be indexed in every conceivable way that someone might think of to identify it: Accession number, title, author, subject, contract, words in an abstract, etc. Searches may be constructed using fields separately or in combination (Fig. 5). Fields may be searched in full or by individual words, by subfields or by masking. Boolean searching of combinations of fields or search results is also possible. The results of all searches can be displayed or printed out by any or all of the fields used in the indexing in any order or in any combination.

The flexibility of the system allows for the addition of a field or subfield if it is discovered that there is a need for one that hadn't been thought of before. Global changes allow for sweeping changes in indexing if that should happen to be deemed necessary. Authority lists and lookup tables are used for controlling subject and descriptive cataloging terms, but fields are also available for the cataloger to use for flights of fancy if desired.

- **Planning for the future**

When the 140-000 report collection currently being scanned is completed, a collection of another 100,000 reports averaging 100 pages each will also be put to disk. This will mean another 10 million pages of images stored on a separate autochanger to be daisy chained to the current one.

Scanning and printing in color will be considered as the technology becomes more affordable

Plans are being made for the providing these images over the campus network to the offices of the Naval Research Laboratory scientists. Future plans include an experiment to examine the possibility of storing

journal articles optically and networking their images to the scientists of the Laboratory. This experiment is part of the Library's on-going efforts to deliver information directly to the researcher's desktop. This capability is currently implemented for text information through the Library's campus-wide information system called InfoNet. InfoNet provides users in their offices and laboratories with a single menu-driven interface through with they can access to CD-ROM databases, library catalogs, laboratory management information databases and Internet resources.

The NRL Library is discussing with other libraries collaborative programs to share images of reports and possibly journals.

To provide full text search capabilities for the most important parts of reports, a Kurtzweil OCR scanner is being brought online to OCR abstracts and the first few pages of stored images and automatically incorporate the resulting ASCII text into the Cuadra Star data base.

Once equipment like the Optical Disk System is in place, possible refinements become obvious and a wish list grows rather quickly - more work stations for patrons and staff, more information in the retrieval data base, faster scanners, faster printers, remote access and all the new products that this rapidly advancing technology is going to offer.

| Home | NRL | Suggest | Search |

webmaster@library.nrl.navy.mil
Updated: 27-NOV-95
Maintained by: Fred Rettenmaier